POLYSEMY IN VERBS: SYSTEMATIC RELATIONS BETWEEN SENSES AND THEIR EFFECT.

Isroilova Jamila

UzSWLU, student of MA department,

Safarova Zarina Giyosovna

Superviser: Teacher of UzSWLU

Annotation.

This article is selected examples of polysemantic verbs in short English sentences. In the method, literalness is distinguished because it satisfies contextual constraints that the nonliteral others all violate. Polysemy is discriminated anomaly in a way that [1] supports Lakoff and Johnson's (1980) view that in polysemantic verbs one entity stands for another whereas in polysemy one entity is viewed as another. Out of the two hundred most frequent polysemantic verbal lexemes of the English language, we have identified **65 verbal** lexemes that fall into the category of multi-seed. Of the total number of tokens in this group, 15 turned out to be six -digit: **become, believe, buy, decide, spend, cause, grow, smile, regard, enable, complete, cost, laugh, arise, define-,** 10 - seven - digit: **need, provide, remain, create, relate, identify, forget, assume, prepare, examine;** 18 - eight - digit: **know, ask, allow, appear, continue, remember, understand, sell, apply, contain, arrive, plan, manage, maintain, admit, refuse, teach, notice,** 11 - nine **-** digit: **want, help, offer, consider, agree, build, involve, stay, affect, determine, fight** and 11 more - ten - digit: **try, bring, pay, hear, win, claim, wish, love, introduce, refer, share.**

Key words: language, style, literature, Harry Potter, occasionalism, formal, polysemy, novel, text.

Annotatsiya.

Ushbu maqola ingliz tilidagi qisqa jumlalardagi polisemantik fe'llarning tanlangan namunalari. Usulda so'zma-so'zlik farqlanadi, chunki u so'zma-so'z bo'lmagan boshqalar buzadigan kontekstli cheklovlarni mavjud. Ko'p ma'nolilik diskriminatsiyalangan anomaliya bo'lib, [1] Lakoff va Jonsonning (1980) ko'p ma'noli fe'llarda bir shaxs boshqasini anglatadi, polisemiyada esa bir shaxs boshqasi sifatida qaraladi, degan fikrni qo'llab-quvvatlaydi. Ingliz tilining ikki yuzta eng tez-tez uchraydigan polisemantik fe'l leksemalaridan biz ko'p urug'li turkumga kiruvchi 65 ta og'zaki leksemani aniqladik. Ushbu guruhdagi tokenlarning umumiy sonidan 15 tasi oltita raqamli bo'lib chiqdi: bo'l, ishon, sotib ol, qaror qil, sarf qil, sabab bo'l, o's,

ISSN 2181-2632 www.involta.uz

tabassum, e'tibor, yoqish, yakunlash, xarajat qilish, kulish, paydo bo'lish, aniqlash-, 10 - etti - raqam: kerak, ta'minla, qol, yaratish, bog'lash, aniqlash, unutish, taxmin qilish, tayyorlash, tekshirish; 18 - sakkiz - raqam: bilish, so'rash, ruxsat berish, paydo bo'lish, davom ettirish, eslash, tushunish, sotish, qo'llash, o'z ichiga olish, kelish, rejalashtirish, boshqarish, saqlash, tan olish, rad etish, o'rgatish, ogohlantirish, 11 - to'qqiz - raqam: xohlayman, yordam berish, taklif qilish, ko'rib chiqish, rozilik berish, qurish, jalb qilish, qolish, ta'sir qilish, aniqlash, jang qilish va yana 11 ta - o'n raqam: harakat qilib ko'ring, keltiring, to'lang, eshiting, yutib oling, da'vo qiling, tilak qiling, seving, tanishtiring, murojaat qiling, baham ko'ring.

Kalit so'zlar: til, uslub, adabiyot, Garri Potter, okkazionalizm, rasmiy, polisemiya, roman, matn.

Introduction; In this paper, we examine different types of relations within sense inventories and give a qualitative analysis of the effects they have on decisions made by the annotators and annotator error. We also discuss some common traps and pitfalls in design of sense inventories. We use the data set developed specifically for the task of annotating sense distinctions dependent predominantly on semantics of the arguments and only to a lesser extent on syntactic frame. 1 Introduction Lexical ambiguity is pervasive in natural language, and its resolution has been used to improve performance of a number of natural language processing (NLP) applications, such as statistical machine translation (Chan et al., 2007; Carpuat and Wu, 2007), cross-language information retrieval and question answering (Resnik, 2006). Sense differentiation for the predicates depends on a number of factors, including syntactic frame, semantics of the arguments and adjuncts, contextual clues from the wider context, text domain identification, etc. Preparing sense-tagged data for training and evaluation of word sense disambiguation (WSD) systems involves two stages: (1) creating a sense inventory and (2) applying it in annotation.

Main chapter; Creating sense inventories for polysemous words is a task that is notoriously difficult to formalize. For polysemous verbs especially, constellations of related meanings make this task even more difficult. In lexicography, “lumping and splitting” senses during dictionary construction – i.e. deciding when to describe a set of usages as a separate sense – is a well-known problem (Hanks and Pustejovsky, c 2008. Licensed under the Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported license (http://creativecommons.org/licenses/by-nc-sa/3.0/). Some rights reserved. 2005; Kilgarriff, 1997). It is often resolved on an adhoc basis, resulting in numerous cases of “overlapping senses”, i.e. instances when the same occurrence may fall under more than one sense category simultaneously. This problem has also been the subject of extensive study in lexical semantics, addressing questions such as when the context selects a distinct sense and when it merely modulates the meaning, what is the regular relationship between related senses, and what compositional processes are involved in sense selection (Pustejovsky, 1995; Cruse, 1995; Apresjan, 1973). A number of syntactic and semantic tests are traditionally applied for sense identification, such as examining synonym series, compatible syntactic environments,

ISSN 2181-2632

coordination tests such as cross-understanding or zeugma test (Cruse, 2000). None of these tests are conclusive and normally a combination of factors is used. At the recent Senseval competitions (Mihalcea et al., 2004; Snyder and Palmer, 2004; Preiss and Yarowsky, 2001), the choice of sense inventories frequently presented problems, spurring the efforts to create coarsergrained sense inventories (Hovy et al., 2006; Palmer et al., 2007; Navigli, 2006). Part of the reason for such difficulties in establishing a set of senses available to a lexical item is that the meaning of a polysemous verb is often determined in composition and depends to the same extent on semantics of the particular arguments as it does on the base meaning of the verb itself. A number of systematic relations often holds between different senses of a polysemous verb.

Depending on the kind of ambiguity involved in each case, some senses are easier to distinguish than others. Sense-/tagged data (e.g. SemCor (Landes et al., 1998), PropBank (Palmer et al., 2005), OntoNotes (Hovy et al., 2006)) typically provides no way to differentiate between sense distinctions motivated by different factors. Treating different disambiguation factors separately would allow one to examine the contribution of each factor, as well as the success of a given algorithm in identifying the corresponding senses. Within the scope of a sentence, syntactic frame and semantics of the arguments are most prominent in sense 33 disambiguation. The latter is often more subtle and hence complex. Our goal in the present study was to target sense distinctions motivated strongly or exclusively by differences in argument semantics. We base the present discussion on the sense-tagged data set we developed for 20 polysemous verbs. We argue below that cases which can not be reliably disambiguated by humans introduce noise into the data and therefore should be kept out, a principle adhered to in the design of this data set. The choice of argument semantics as the target disambiguation factor was motivated by several considerations. In automatic sense detection systems, argument semantics is often represented using external resources such as thesauri or shallow ontologies. Sense induction systems using distributional information often do not take into account the possible implications of induced word clusters for sense disambiguation. Our goal was to analyze differences in argument semantics that contribute to disambiguation. In this paper, we discuss different kinds of systematic relations observed between senses of polysemous predicates and examine the effects they have on decisions made by the annotators. We also examine sense inventories for other factors that influence inter-annotator agreement rates and lead to annotation error. In Section 2, we discuss some of the factors that influence compilation of sense inventories and the methodology involved. In Section 3, we describe briefly the data set and the annotation task. In Sections 4 and 5, we discuss the relations observed between different senses within sense inventories in our data set, their effect on decisions made by the annotators, and the related annotation errors. 2 Defining A Sense Inventory Several current resource-oriented projects undertake to formalize the procedure of identifying a word sense. FrameNet (Ruppenhofer et al., 2006) attempts to organize lexical information in terms of script-like semantic frames, with semantic and syntactic combinatorial possibilities specified for each frame-evoking lexical unit (word/sense pairing). Semantics of the arguments is represented by Fillmore’s case roles (frame elements) which are derived on ad-hoc basis for each frame. In

ISSN 2181-2632 www.involta.uz

OntoNotes project, annotators use small-scale corpus analysis to create sense inventories derived by grouping together WordNet senses. The procedure is restricted to maintain 90% inter-annotator agreement (Hovy et al., 2006).

Corpus Pattern Analysis (CPA) (Hanks and Pustejovsky, 2005; Pustejovsky et al., 2004) attempts to catalog prototypical norms of usage for individual words, specifying them in terms of context patterns. As a corpus analysis technique, CPA has its origins in the analysis of large corpora for lexicographic purposes, of the kind that was used for compiling the Cobuild dictionary (Sinclair and Hanks, 1987). Each pattern gives a combination of surface textual clues and argument specifications. A lexicographer creates a set of patterns by sorting a concordance for the target predicate according to the context features. In the present study, we use a modification of the CPA technique in the way explained in Section 3. In CPA, syntactic and textual clues include argument structure and minor syntactic categories such as locatives and adjuncts; collocates from wider context; subphrasal cues such as genitives, partitives, bare plural/determiner, infinitivals, negatives, etc. Semantics of the arguments is represented either through a set of shallow semantic types corresponding to basic semantic features (e.g. Person, Location, PhysObj, Abstract, Event, etc.) or extensionally through lexical sets, which are effectively collections of lexical items.1 Several CPA patterns may correspond to a single sense. The patterns vary in syntactic structure or the encoding of semantic roles relative to the described event. For example, for the verb treat, DOCTOR treating PATIENT and DOCTOR treating DISEASE both correspond to the medical sense of treat. Knowing which semantic role is expressed by a particular argument is often useful for performing inference. For instance, treating a disease eliminates the disease, but not the patient. In the present annotation task, each pattern is viewed as sense in construction and labeled as a separate sense. In the rest of the paper, we will use the term “sense” to refer also to such microsenses. For the cases where sense differentiation depends strongly on differences in semantics of the arguments, several factors further complicate creating a sense inventory. Prototypicality as a general principle of category organization seems to play an important role in defining both the boundaries of senses and the corresponding argument groupings. The same sense of the predicate is often activated by a number of semantically diverse arguments. Such argument sets are frequently organized around a core of typical members that are a “good fit” with respect to semantic requirements of the corresponding sense of the target. The relevant semantic feature is prominent for them, while other, more peripheral members of the argument set, merely allow the relevant interpretation (see Rumshisky (2008) for discussion). For example, the verb absorb has a sense involving absorbing a substance, and the typical members of the corresponding argument set would be actual substances, such as oil, oxygen, water, air, salt, etc. But goodness, dirt, flavor, moisture would also activate the same sense. Each decision to split a sense and make another category is to a certain extent an arbitrary decision. For example, for the verb absorb, one can separate absorbing a substance (oil, oxygen, water, air, salt) from absorbing energy (radiation, heat, sound, energy). The latter sense may or may not be separated from absorb1 See Rumshisky et al. (2006) and Pustejovsky et al. (2004) for more detail. 34 ing impact (blow, shock, stress). But

ISSN 2181-2632 www.involta.uz

it is a marked continuum, i.e. certain points in the continuum are more prominent, with necessity of a given concept reflected in the frequency of use. When several senses are postulated based on argument distinctions, there are almost always boundary cases that can be seen to belong to both categories. Consider, for example, two senses defined for the verb launch and the corresponding direct objects in (1): (1) a. Physically propel an object into the air or water missile, rocket, torpedo, satellite, shuttle, craft b. Begin or initiate an endeavor campaign, initiative, investigation, expedition, drive, competition, crusade, attack, assault, inquiry The senses seem to be very clearly separated, yet examples like launch a ship clearly fall on the boundary: while ships are physical objects propelled into water, launching a ship can be virtually synonymous with launching an expedition. Similarly, for the verb conclude, two senses below which are linked to nominal complements are clearly separated: (2) a. finish meeting, debate, investigation, visit, tour, discussion; letter, chapter, novel b. reach an agreement treaty, agreement, deal, contract, truce, alliance, ceasefire, sale However, conclude negotiations is clearly a boundary case where both interpretations are equally possible (negotiations may be concluded without reaching an agreement). In fact, the two annotators chose different senses for this example:2 (3)

We were able to operate under a lease agreement until purchase negotiations were concluded. annoA: finish annoB: reach an agreement In many cases, postulating a separate sense for a coherent set of nominal complements is not justified, as there are regular semantic processes that allow the complements to satisfy selectional requirements of the verb. For example, the verb conclude, in the finish sense accepts EVENT complements. Therefore, nouns such as letter, chapter, novel in (2) must be coerced into events corresponding to the activity that typically brings them about, that is, re-interpreted as events of writing (their Agentive quale, cf. Pustejovsky (1995)). Similarly, the verb deny in the first sense (state or maintain that something is untrue) accepts PROPOSITION complements: (4) a. state or maintain that something is untrue allegations, reports, rumour; significance, importance, difference; attack, assault, involvement b. refuse to grant something access, visa, approval, funding, license 2All examples are taken from the annotated data set. In some cases, sentence structure was slightly modified for brevity. Event nouns such as attack and assault are coerced into a propositional reading, as are relational nouns such as significance and importance. Interestingly, as we have noted before (Rumshisky et al., 2006), each predicate imposes its own gradation with respect to prototypicality of elements of the argument set. As a result, even though basic semantic types such as PHYSOBJ, ANIMATE, EVENT, are used uniformly by many predicates, argument sets, while semantically similar, typically differ between predicates. For example, fall in the subject position and cut in the direct object position select for things that can be decreased: (5) a. cut (dobj): reduce or lessen price, inflation, profits, cost, emission, spending, deficit, wages overhead, production, consumption, fees, staff b. fall (subj): decrease price, inflation, profits, attendance, turnover, temperature, membership, import, demand, level While there is a clear commonality between these argument sets, the overlap is only partial. To give another example, consider INFORMATION-selecting predicates explain (subj), grasp (dobj) and know (dobj). The nouns book and note occur in the subject position of explain; answer

ISSN 2181-2632 www.involta.uz

occurs both as the subject of explain and direct object of know; however, grasp accepts neither of these nouns as direct object. Thus, the actual selectional behavior of the predicates does not seem to be well described in terms of a fixed set of types, which is what is typically assumed by many ontologies used in automatic WSD.

Conclusion; We were interested specifically in those cases where disambiguation needs to be made without relying on syntactic frame, and the main source of disambiguation is semantics of the arguments. Such cases are harder to identify formally in the development of sense inventories and harder for the annotators to determine. For example, phrasal verbs or idiomatic constructions that help identify a particular sense were intentionally excluded from our data set. Thus, for the verb cut, one of the senses involves cutting out a shape or a form (e.g. cut a suit), but the sentences with the corresponding phrasal form cut out were thrown out. Even so, syntactic clues that contribute to disambiguation in some cases overrule the interpretation suggested by the argument. For example, for the verb deny, in deny the attack, the direct object strongly suggests a propositional interpretation for deny (that the attack didn’t happen). However, the use of ditransitive construction (indicated in the example below by the past participle) overrules this interpretation, and we get the refuse to grant sense: (6) Astorre, denied his attack, had stayed in camp, uneasily brooding. 35 In fact, during the actual annotation, one of the annotators did not recognize the use of past participle, and erroneously assigned the state or maintain something to be untrue sense to this sentence. 3.1 Data set The data set was developed using the British National Corpus (BNC), which is more balanced than the more commonly annotated Wall Street Journal data. We selected 20 polysemous verbs with sense distinctions that were judged to depend for disambiguation on semantics of the argument in several argument positions, including direct object (dobj), subject (subj), or indirect object within a prepositional phrase governed by with (iobj with): dobj: absorb, acquire, admit, assume, claim, conclude, cut, deny, dictate, drive, edit, enjoy, fire, grasp, know, launch subj: explain, fall, lead iobj with: meet We used the Sketch Engine (Kilgarriff et al., 2004) both to select the verbs and to aid the creation of the sense inventories. The Sketch Engine is a lexicographic tool that lists collocates that co-occur with a given target word in the specified grammatical relation. The collocates are sorted by their association score with the target. A set of senses was created for each verb using a modification of the CPA technique (Pustejovsky et al., 2004). A set of complements was examined in the Sketch Engine. If a clear division was observed between semantically different groups of collocates in a certain argument position, the verb was selected. For semantically distinct groups of collocates, a separate sense was added to the sense inventory for the target. For example, for the verb acquire, a separate sense was added for each of the following sets of direct objects: (7) a.

Used literatura.

1. Halliday, M. A. K. Til funktsiyalaridagi tadqiqotlar London: Edvard Arnold. 1973. - 176-bet

ISSN 2181-2632 www.involta.uz

2. Guver, D.L. The Merosors in Language and Style Lanham, MD: University Press of America. 1999.-B.350-52

3. Paul Simpson Stilistika: talabalar uchun resources kitobi. New York: Roudledj, 2004. - 88-bet

4. https://www.futurelearn.com>

ISSN 2181-2632 www.involta.uz